



OP-2641

PATENT

Docket No.: 36409-00900

**IN THE UNITED STATES PATENT AND TRADEMARK OFFICE**

Applicant(s) : Masayuki Yamada  
Serial No. : 09/821,671 Group Art Unit: 2641  
Filed : March 29, 2001 Examiner: TBA  
For : ***SPEECH SYNTHESIZING METHOD AND APPARATUS***

**CLAIM TO CONVENTION PRIORITY**

Box : No Fee  
COMMISSIONER OF PATENTS  
Washington, D.C. 20231

Sir:

In the matter of the above-identified application and under the provisions of 35 U.S.C. § 119 and 37 C.F.R. § 1.55 applicant(s) claim(s) the benefit of the following prior application:

Application filed in : Japan  
In the name of : Canon Kabushiki Kaisha  
Serial No. : 2000-099531  
Filing Date : March 31, 2000

1. ☒ Pursuant to the Claim to Priority, applicant(s) submit(s) a duly certified copy of said foreign application.
2. ☐ A duly certified copy of said foreign application is in the file of application Serial No. \_\_\_\_\_, filed \_\_\_\_\_.

Respectfully submitted,  
Milbank, Tweed, Hadley & McCloy, L.L.P.

Chris L. Holm  
Reg. No.: 39,227

June 8, 2001

Milbank, Tweed, Hadley & McCloy LLP  
1 Chase Manhattan Plaza  
New York, NY 10005-1413  
(212) 530-5000 / (212) 530-5219 (facsimile)

NY2:#4417262

RECEIVED  
JUN 14 2001  
Technology Center 2600

RECEIVED (MB)  
JUN 28 2001  
Technology Center 2600

**CERTIFICATE OF MAILING BY "EXPRESS MAIL" (37 CFR 1.10)**

Applicant(s): Masayuki Yamada

Docket No.

36409-00900

Serial No.

09/821,671

Filing Date

March 29, 2001

Examiner

TBA

Group Art Unit

2641

Invention:

SPEECH SYNTHESIZING METHOD AND APPARATUS

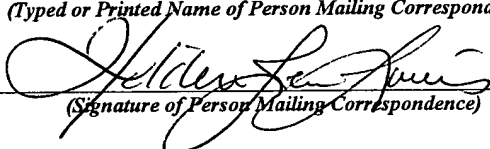
RECEIVED  
JUN 14 2001  
Technology Center 2600

I hereby certify that the following correspondence:

Claim to Convention Priority Transmittal, Japanese Patent application No. 2000-099531 ,Acknowledgment Postcard and Certificate of Mailing By Express Mail No. EL705884378US.

*(Identify type of correspondence)*

is being deposited with the United States Postal Service "Express Mail Post Office to Addressee" service under 37 CFR 1.10 in an envelope addressed to: The Assistant Commissioner for Patents, Washington, D.C. 20231

June 8, 2001*(Date)*Hildere Jean-Louis*(Typed or Printed Name of Person Mailing Correspondence)*  
*(Signature of Person Mailing Correspondence)*EL705884378US*("Express Mail" Mailing Label Number)*

Note: Each paper must have its own certificate of mailing.

RECEIVED

JUN 28 2001

Technology Center 2600



日本国特許庁  
JAPAN PATENT OFFICE

CFM 2177 U.S.

RECEIVED  
JUN 14 2001  
Technology Center 2600

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office

出願年月日  
Date of Application:

2000年 3月31日

出願番号  
Application Number:

特願2000-099531

出願人  
Applicant(s):

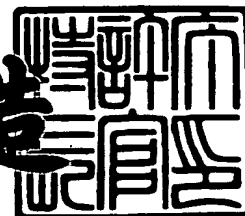
キヤノン株式会社

RECEIVED  
JUN 28 2001  
Technology Center 2600

2001年 4月20日

特許庁長官  
Commissioner,  
Japan Patent Office

及川耕造



出証番号 出証特2001-3033158

【書類名】 特許願

【整理番号】 4172016

【提出日】 平成12年 3月31日

【あて先】 特許庁長官殿

【国際特許分類】 G10L 5/04

【発明の名称】 音声合成方法、装置および記憶媒体

【請求項の数】 21

【発明者】

    【住所又は居所】 東京都大田区下丸子3丁目30番2号 キヤノン株式会社  
社内

    【氏名】 山田 雅章

【特許出願人】

    【識別番号】 000001007

    【氏名又は名称】 キヤノン株式会社

【代理人】

    【識別番号】 100076428

    【弁理士】

    【氏名又は名称】 大塚 康德

    【電話番号】 03-5276-3241

【選任した代理人】

    【識別番号】 100101306

    【弁理士】

    【氏名又は名称】 丸山 幸雄

    【電話番号】 03-5276-3241

【選任した代理人】

    【識別番号】 100115071

    【弁理士】

    【氏名又は名称】 大塚 康弘

    【電話番号】 03-5276-3241

【手数料の表示】

【予納台帳番号】 003458

【納付金額】 21,000円

【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【包括委任状番号】 0001010

【プルーフの要否】 要

【書類名】 明細書

【発明の名称】 音声合成方法、装置および記憶媒体

【特許請求の範囲】

【請求項 1】 所定単位の音声素片を音素境界で分割して部分素片を取得する分割工程と、

前記分割工程で得られた部分素片毎にパワー値を推定する推定工程と、

前記部分素片の各々のパワー値を前記推定工程で推定されたパワー値に基づいて変更する変更工程と、

前記変更工程で変更された部分素片を用いて合成音声を生成する生成工程とを備えることを特徴とする音声合成方法。

【請求項 2】 前記変更工程は、前記部分素片の各々について、

対応するパワー基準値を取得し、

前記推定工程で推定されたパワー値および前記取得されたパワー基準値に基づいて振幅変更倍率を計算し、

前記計算された振幅変更倍率に従って当該部分素片の振幅を変更することにより前記推定されたパワー値への変更を行う

ことを特徴とする請求項 1 に記載の音声合成方法。

【請求項 3】 前記変更工程は、前記推定工程で推定したパワー値を  $p$ 、前記取得したパワー基準値を  $q$  とした場合に、

$$s = (p / q)^{1/2}$$

によって求まる  $s$  を振幅変更倍率として、当該部分素片の振幅値を変更する

ことを特徴とする請求項 2 に記載の音声合成方法。

【請求項 4】 前記推定工程は、

前記部分素片の各々について有声音であるか無声音であるかを判定する判定工程を更に備え、

有声音であると判定された場合には有声音素片用のパラメータ値でパワー値を推定し、無声音であると判定された場合には無声音素片用のパラメータ値でパワー値を推定する

ことを特徴とする請求項 1 に記載の音声合成方法。

【請求項5】 前記推定工程は、  
各部分素片毎にパワー推定用要因を獲得する獲得工程を更に備え、  
前記判定工程の判定結果に応じて、前記獲得されたパワー推定用要因に対応するパラメータ値を取得してパワー値を推定する  
ことを特徴とする請求項4に記載の音声合成方法。

【請求項6】 前記パワー推定用要因は、当該部分素片の音素種別、当該部分素片の合成対象語におけるモーラ位置、合成対象語のモーラ数及びアクセント型のいずれかを含む  
ことを特徴とする請求項5に記載の音声合成方法。

【請求項7】 前記獲得工程は、前記判定工程によって有声音であると判定された場合は有声音用のパワー推定用要因を獲得し、無声音であると判定された場合は無声音用のパワー推定用要因を獲得する  
ことを特徴とする請求項6に記載の音声合成方法。

【請求項8】 無声音の部分素片に対応するパワー基準値が相対的に大きめの値に設定されている  
ことを特徴とする請求項4乃至7のいずれかに記載の音声合成方法。

【請求項9】 前記音声合成単位がCV/V Cである  
ことを特徴とする請求項1乃至8のいずれかに記載の音声合成方法。

【請求項10】 前記音声合成単位がVCVである  
ことを特徴とする請求項1乃至8のいずれかに記載の音声合成方法。

【請求項11】 所定単位の音声素片を音素境界で分割して部分素片を取得する分割手段と、  
前記分割手段で得られた部分素片毎にパワー値を推定する推定手段と、  
前記部分素片の各々のパワー値を前記推定手段で推定されたパワー値に基づいて変更する変更手段と、  
前記変更手段で変更された部分素片を用いて合成音声を生成する生成手段とを備えることを特徴とする音声合成装置。

【請求項12】 前記変更手段は、前記部分素片の各々について、  
対応するパワー基準値を取得し、

前記推定手段で推定されたパワー値および前記取得されたパワー基準値に基づいて振幅変更倍率を計算し、

前記計算された振幅変更倍率に従って当該部分素片の振幅を変更することにより前記推定されたパワー値への変更を行う

ことを特徴とする請求項 11 に記載の音声合成装置。

【請求項 13】 前記変更手段は、前記推定手段で推定したパワー値を  $p$ 、前記取得したパワー基準値を  $q$  とした場合に、

$$s = (p / q)^{1/2}$$

によって求まる  $s$  を振幅変更倍率として、当該部分素片の振幅値を変更する

ことを特徴とする請求項 12 に記載の音声合成装置。

【請求項 14】 前記推定手段は、

前記部分素片の各々について有声音であるか無声音であるかを判定する判定手段を更に備え、

有声音であると判定された場合には有声音素片用のパラメータ値でパワー値を推定し、無声音であると判定された場合には無声音素片用のパラメータ値でパワー値を推定する

ことを特徴とする請求項 11 に記載の音声合成装置。

【請求項 15】 前記推定手段は、

各部分素片毎にパワー推定用要因を獲得する獲得手段を更に備え、

前記判定手段の判定結果に応じて、前記獲得されたパワー推定用要因に対応するパラメータ値を取得してパワー値を推定する

ことを特徴とする請求項 14 に記載の音声合成装置。

【請求項 16】 前記パワー推定用要因は、当該部分素片の音素種別、当該部分素片の合成対象語におけるモーラ位置、合成対象語のモーラ数及びアクセント型のいずれかを含む

ことを特徴とする請求項 15 に記載の音声合成装置。

【請求項 17】 前記獲得手段は、前記判定手段によって有声音であると判定された場合は有声音用のパワー推定用要因を獲得し、無声音であると判定された場合は無声音用のパワー推定用要因を獲得する



ことを特徴とする請求項 16 に記載の音声合成装置。

【請求項 18】 無声音の部分素片に対応するパワー基準値が相対的に大きな値に設定されている

ことを特徴とする請求項 14 乃至 17 のいずれかに記載の音声合成装置。

【請求項 19】 前記音声合成単位が CV/VC である

ことを特徴とする請求項 11 乃至 18 のいずれかに記載の音声合成装置。

【請求項 20】 前記音声合成単位が VCV である

ことを特徴とする請求項 11 乃至 18 のいずれかに記載の音声合成装置。

【請求項 21】 請求項 1 乃至 10 のいずれかに記載の方法をコンピュータに実現させるための制御プログラムを格納する記憶媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は、音声合成方法および装置に関し、特に音声合成時における合成音声のパワー制御に関わるものである。

【0002】

【従来の技術】

所望の合成音声を得るための音声合成方法には、音素や CV・VC あるいは VCV 等の音韻を単位とした音声素片を編集、接続して合成音声を生成する方法が知られている。図 10 は音声素片単位である CV・VC、VCV (C: 子音, V: 母音) を説明する図である。図 10 に示されるように、CV・VC は各音素内に素片境界を置いた単位であり、VCV は母音内に素片境界を置いた単位である。

【0003】

【発明が解決しようとする課題】

図 11 は、1 音声素片の継続時間長や基本周波数を変更する方法の一例を模式的に示した図である。図 11 の上段に示す 1 音声素片の音声波形は、中段に示す複数個の窓関数によって複数個の微細素片に分割される。このとき、有声音部 (音声波形の後半部にある有声音の領域) では、原音声のピッチ間隔に同期した時

間幅を有する窓関数を用いる。一方、無声音部（音声波形の前半部にある無声音の領域）では、適当な時間幅（一般には、有声音部の窓関数よりも長い時間幅を有する）の窓関数を用いる。

#### 【 0 0 0 4 】

このようにして得た複数個の微細素片を繰り返したり、間引いたり、間隔を変更したりすることによって、合成音声の継続時間長や基本周波数を変更することができる。例えば、合成音声の継続時間長を短縮する場合には、微細素片の間引けばよく、合成音声の継続時間長を伸長する場合には、微細素片を繰り返せばよい。また、合成音声の基本周波数を上げる場合には、有声音部の微細素片の間隔を詰めればよく、合成音声の基本周波数を下げる場合には、有声音部の微細素片の間隔を広げればよい。このような繰り返し、間引き、間隔変更を施して得た複数個の微細素片を重畳することにより、所望の継続時間長、基本周波数を有する合成音声を得ることができる。

#### 【 0 0 0 5 】

また、このような合成音声に対するパワー制御は以下のように行われる。すなわち、所望の平均パワーを持つ合成音声は、音声素片の平均パワーの推定値  $p_0$ （目標とする平均パワーに対応する）と上記手順によって得られた合成音声の平均パワー  $p$  とを求め、上記手順によって得られた合成音声に  $(p/p_0)^{1/2}$  を乗ずることにより得られる。つまり、1 音声素片単位にパワー制御を実行する。

#### 【 0 0 0 6 】

しかしながら、上記のパワー制御方法には以下の問題点がある。

#### 【 0 0 0 7 】

まず第一の問題点としてパワー制御の単位と音声素片の単位とのミスマッチの問題がある。

安定したパワー制御を行うためには、ある程度長い時間を単位としてパワー制御を行う必要がある。また、パワー制御単位内では、パワー変動が少ないことも必要である。これらの条件を満たすパワー制御の単位には、音素あるいは音素に類した単位がある。しかしながら、上述した CV・VC あるいは VCV といった単位では、変動の激しい音素境界を素片内部に持ったため、素片内部でのパワー

変動が大きくなり、パワー制御の単位としては不適當である。

【0008】

有声音部と無声音部ではパワーの値に大きな差がある。原則的には音素種別から有声音／無声音の別は一意に定まるため、音素毎にパワーの平均値を推定すれば、この差が問題になることはないことになる。しかし、詳細に調べると、音素種別と有声音／無声音の関係には例外があり、ミスマッチが生じることがある。また、音素境界と有声音／無声音境界が数msecから十数msec程度ずれる場合もある。これは、音素種別および音素境界が、主に声道形状によって定められるものであるのに対し、有声音／無声音は声帯振動の有無によるためである。

【0009】

本発明は、上記の問題に鑑みてなされたものであり、その目的は、音声素片内のパワー変動が大きくなるような音韻単位を波形編集の単位としても適切なパワー制御を行うことを可能にすることにある。

【0010】

【課題を解決するための手段】

上記の目的を達成するための本発明の一態様による音声合成方法は例えば以下の構成を備える。すなわち、

所定単位の音声素片を音素境界で分割して部分素片を取得する分割工程と、

前記分割工程で得られた部分素片毎にパワー値を推定する推定工程と、

前記部分素片の各々のパワー値を前記推定工程で推定されたパワー値に基づいて変更する変更工程と、

前記変更工程で変更された部分素片を用いて合成音声を生成する生成工程とを備える。

【0011】

また、上記の目的を達成するための本発明の他の態様によれば、上記構成の音声合成装置において、

所定単位の音声素片を音素境界で分割して部分素片を取得する分割手段と、

前記分割手段で得られた部分素片毎にパワー値を推定する推定手段と、

前記部分素片の各々のパワー値を前記推定手段で推定されたパワー値に基づい

て変更する変更手段と、

前記変更手段で変更された部分素片を用いて合成音声を生成する生成手段とを備える。

【 0 0 1 2 】

更に、本発明の他の態様によれば、上記の音声合成方法をコンピュータに実現させるための制御プログラムを格納した記憶媒体が提供される。

【 0 0 1 3 】

【発明の実施の形態】

以下、添付の図面を参照して本発明の好適な実施形態を説明する。

【 0 0 1 4 】

〔第 1 の実施形態〕

図 1 は本実施形態による音声合成装置のハードウェア構成を示すブロック図である。図 1 において、11 は数値演算・制御等の処理を行なう中央処理装置であり、図 2 のフローチャートで後述する制御を実現する。12 は RAM、ROM 等の記憶装置であり、図 2 のフローチャートで後述する制御を中央処理装置 11 に実現させるために必要な制御プログラムや一時的データが格納される。13 はディスク装置等の外部記憶装置であり、本実施形態の音声合成処理を制御する制御プログラムやユーザの操作を受けるためのグラフィカルユーザインタフェースを制御する制御プログラムを保持する。

【 0 0 1 5 】

14 は表示器、スピーカ等からなる出力装置であり、合成された音声はスピーカから出力される。また、表示器には、ユーザの操作を受け付けるグラフィカルユーザインタフェースを表示する。このグラフィカルユーザインタフェースは、中央処理装置 11 によって制御される。ただし、本発明は他の装置やプログラムに対して合成音声を出力するべく組み込むことも可能であり、この場合の出力は他の装置或いはプログラムの入力となる。15 はキーボード等の入力装置であり、ユーザの操作を所定の制御コマンドに変換して中央処理装置 11 に供給する。中央処理装置 11 は、この制御コマンドの内容に応じて、音声合成の対象となるテキスト（日本語や他の言語からなる）を指定し、そのテキストを音声合成ユニ

ット17に供給する。ただし、本発明は他の装置やプログラムの一部として組み込まれることも可能であり、この場合の入力は他の装置やプログラムを通じて間接的に行われることになる。16は内部バスであり、図1で示された上述の各構成を接続する。17は音声合成ユニットである。音声合成ユニット17は、素片辞書18を用いて、入力したテキストから音声合成する。但し、素片辞書18は、外部記憶装置13が保持するように構成してもよい。

## 【0016】

以上のハードウェア構成を備えた本実施形態の音声合成ユニット17の動作を以下に説明する。

## 【0017】

図2は本実施形態による音声合成ユニット17の手順を示すフローチャートである。まず、ステップS1において、入力したテキストに対して言語解析と音響処理を施し、そのテキストを表す音韻系列とその音韻系列の音韻情報（モーラ数、モーラ位置、アクセント型等）とを生成する。次に、1音韻単位（合成単位ともいう）の音声素片を表す音声波形データを素片辞書18から読み出す。ここで、音韻の単位は、CV・VC、VCV等の音素境界を含む単位である。次に、ステップS2において、ステップS1において取得した音声素片を音素境界を境界として分割する。ステップS2の分割によって得られた素片を部分素片 $u_i$ と呼ぶことにする。例えば、音声素片がVCVであるならば3個、CV・VCならば2個の部分素片に分割されることになる。次に、ステップS3においてループカウンタ $i$ を0に初期化する。

## 【0018】

続くステップS4において、部分素片 $u_i$ のパワー推定に必要な推定要因を取得する。本例では、図3に示されるように、部分素片 $u_i$ の音素種別、合成対象語のアクセント型及びモーラ数、合成対象語中における部分素片 $u_i$ の置かれる位置（モーラ位置に相当する）等が推定要因として用いられる。これらの推定要因は、ステップS1で得た音韻情報に含まれる。ステップS5では、部分素片 $u_i$ が有声音の素片か無声音の素片かを判定するための情報（図4）を取得する。すなわち、ステップS1で取得した音声素片に対応する素片IDとその音声素片

の部分素片番号（ループカウンタ  $i$  が対応する）とから対応する有声音無声音フラグを取得する。図4に示す情報は、素片辞書18が保持する。

【0019】

次に、ステップS6において、ステップS5で得た有声音無声音フラグに基づいて、部分素片  $u_i$  が有声音の素片か無声音の素片かを判断して処理を分岐する。すなわち、部分素片  $u_i$  が有声音の場合はステップS7に処理を移し、 $u_i$  が無声音の場合にはステップS9に処理を移す。

【0020】

ステップS7では、上述のステップS4で得られた各推定要因に基づいて、有声音パワー推定用のパラメータ値を取得する。例えば、数量化I類による推定を行う場合には、有声音パワー推定用に学習された数量化I類の係数表（図5）から、ステップS4で得られた推定要因に対応するパラメータ値を取得する。そして、ステップS8において、ステップS7で得られたパラメータ値に基づいて合成音ターゲットとなるパワー  $p$  を推定し、ステップS11に移る。尚、図5に示す情報は、素片辞書18が保持する。

【0021】

一方、部分素片  $u_i$  が無声音であった場合は、ステップS9において、上述のステップS4で得られた各推定要因に基づいて、無声音パワー推定用のパラメータ値を取得する。例えば、数量化I類による推定を行う場合には、無声音パワー推定用に学習された数量化I類の係数表（図6）から、ステップS4で得られた推定要因に対応するパラメータ値を取得する。そして、ステップS10において、ステップS9で得られたパラメータ値に基づいて合成音ターゲットとなるパワー  $p$  を推定し、ステップS11に移る。尚、図5に示す情報は、素片辞書18が保持する。

【0022】

ステップS11では、素片辞書18に記憶された部分素片  $u_i$  に対応するパワー基準値  $q$  を取得する。次に、ステップS12において、ステップS8あるいはステップS10で推定された推定値  $p$  と、ステップS11で取得されたパワー基準値  $q$  とから振幅変更倍率  $s_i$  を計算する。ここで、 $p$ 、 $q$  とともにパワー次元の

値ならば、

$$s_i = (p/q)^{1/2} \text{となる。}$$

【0023】

その後、ステップS13において、ループカウンタ*i*の値に1を加える。次に、ステップS14において、ループカウンタ*i*が1音素単位の部分素片の総数に等しいかどうかを判定し、等しくない場合にはステップS4に戻り、次の部分素片に対して上述の処理を行う。そして、ループカウンタ*i*が、部分素片の総数に等しい場合には、ステップS15に処理を移す。ステップS15では、ステップS12で求めた振幅変更倍率 $s_i$ を用いて、各音声素片の部分素片毎にパワー制御を行う。更に、その他の韻律情報（継続時間長や基本周波数）を用いて、各音声波形に対して波形編集操作を行う。更に、これらの音声素片を接続することにより入力したテキストに対応した合成音を得る。この合成音は、出力装置14のスピーカから出力される。ステップS15は、PSOLA（Pitch-Synchronous Overlap Add method「ピッチ同期波形重畳法」）を用いて、各音声素片の波形編集を行う。

【0024】

以上のように、第1の実施形態によれば、1つ以上の音素境界を含む音声素片を音素境界で分割して部分素片を取得し、各部分素片が有声音か無声音かに応じてパワー推定値を計算することができる。これにより、CV・VCやVCVのように音声素片内のパワー変動が大きくなるような音韻単位を波形編集の単位としても適切なパワー制御を行うことができ、高品位な合成音声を生成することができる。

【0025】

#### 〔第2の実施形態〕

第1の実施形態において、パワー推定用の要因は有声音／無声音に関わらず同じものとしたが、有声音／無声音によってパワー推定用の要因を分けることも可能である。図7は第2の実施形態による音声合成処理の手順を説明するフローチャートである。図7において第1の実施形態（図2）と同様の処理を行うステップには同一のステップ番号を付し、ここではそれらの説明を省略する。

## 【0026】

第1の実施形態ではステップS4において有声音／無声音に関わらず同一のパワー推定用の要因を取得したが、第2の実施形態では、ステップS4を廃し、ステップS16とステップS17で有声音・無声音のそれぞれに応じたパワー推定要因を取得する。すなわち、ステップS6において部分素片 $u_i$ が有声音であると判定された場合は、ステップS16において有声音用のパワー推定要因を取得し、ステップS7ではこの有声音用のパワー推定要因に対応するパラメータ値を図5のテーブルから取得する。一方、ステップS6において、部分素片 $u_i$ が無声音であると判定された場合は、ステップS17において無声音用のパワー推定要因を取得し、ステップS9ではこの無声音用のパワー推定要因に対応するパラメータ値を図6のテーブルから取得する。

## 【0027】

以上のように、第2の実施形態によれば、有声音部と無声音部とで更に適切なパワー制御を行うことができる。

## 【0028】

## 〔第3の実施形態〕

第1及び第2の実施形態において、部分素片のパワー基準値 $q$ として任意の値を用いることが可能であるが、その一例として音素パワーを用いることができる。本実施形態では、部分素片のパワー基準値 $q$ として音素パワーを用いる場合における素片辞書作成の処理を説明する。図8は、音声合成ユニット17における素片辞書作成の処理手順を説明するフローチャートである。また、図9は図8のフローチャートによる素片辞書作成処理を説明する図である。

## 【0029】

まず、ステップS21において素片辞書18に登録すべき発声（図9の（a）、（b））を取得する。次に、ステップS22において、上記ステップS21で取得した発声を音素に分割する（図9の（c））。次に、ステップS23においてループカウンタ $i$ を0に初期化する。

## 【0030】

ステップS24において、 $i$ 番目の音素 $u_i$ の有声音／無声音の別を判定する



。そして、ステップ S 2 5 において、ステップ S 2 4 の判定結果に基づいて処理を分岐する。すなわち、ステップ S 2 4 において当該音素  $u_i$  が有声音であると判定されたならばステップ S 2 6 に処理を移し、無声音であると判定されたならばステップ S 2 8 に処理を移す。

#### 【0031】

ステップ S 2 6 では、 $i$  番目の音素の有声音部の平均パワーを計算する。そしてステップ S 2 7 において、ステップ S 2 6 で計算された有声音部平均パワーをパワー基準値として設定し、ステップ S 3 0 に処理を移す。一方、ステップ S 2 8 では、 $i$  番目の音素の無声音部の平均パワーを計算する。そして、ステップ S 2 9 において、ステップ S 2 8 で計算された無声音部平均パワーをパワー基準値として設定し、ステップ S 3 0 に処理を移す。

#### 【0032】

ステップ S 3 0 では、ループカウンタ  $i$  の値に 1 を加える。そして、ステップ S 3 1 において、ループカウンタ  $i$  が音素の総数に等しいか判定し、等しくない場合には、次の音素について上述の処理を繰り返すべく、ステップ S 2 4 に処理を戻す。一方、ステップ S 3 1 でループカウンタ  $i$  が音素数に等しいと判定された場合は、本処理を終了する。以上の処理により、図 9 の (d) の如く各音素の有声音・無声音が判定され、図 9 の (e) に示す如く音素パワー基準値が設定される。

#### 【0033】

そして、上述のステップ S 1 1 においては、例えば、CV・VC 単位の音声素片「t.a」を部分素片 /t/ と /a/ に分割した場合には、/t/ のパワー基準値  $q$  として「893」が、/a/ のパワー基準値  $q$  として「2473」が用いられることになる（図 9 の (e) ～ (g)）。

#### 【0034】

なお、上記第 3 の実施形態において、ステップ S 2 9 において、無声音部の平均パワーに 1 より大きな値を乗じた値をパワー基準値とすることにより、合成時の無声音のパワーを更に抑える効果が得られる。これは、上述したステップ S 1 2 における変更倍率の値が小さくなるからである。

## 【 0 0 3 5 】

なお、本発明の目的は、前述した実施形態の機能を実現するソフトウェアのプログラムコードを記録した記憶媒体（または記録媒体）を、システムあるいは装置に供給し、そのシステムあるいは装置のコンピュータ（またはCPUやMPU）が記憶媒体に格納されたプログラムコードを読み出し実行することによっても、達成されることは言うまでもない。この場合、記憶媒体から読み出されたプログラムコード自体が前述した実施形態の機能を実現することになり、そのプログラムコードを記憶した記憶媒体は本発明を構成することになる。また、コンピュータが読み出したプログラムコードを実行することにより、前述した実施形態の機能が実現されるだけでなく、そのプログラムコードの指示に基づき、コンピュータ上で稼働しているオペレーティングシステム（OS）などが実際の処理の一部または全部を行い、その処理によって前述した実施形態の機能が実現される場合も含まれることは言うまでもない。

## 【 0 0 3 6 】

さらに、記憶媒体から読み出されたプログラムコードが、コンピュータに挿入された機能拡張カードやコンピュータに接続された機能拡張ユニットに備わるメモリに書込まれた後、そのプログラムコードの指示に基づき、その機能拡張カードや機能拡張ユニットに備わるCPUなどが実際の処理の一部または全部を行い、その処理によって前述した実施形態の機能が実現される場合も含まれることは言うまでもない。

## 【 0 0 3 7 】

## 【発明の効果】

以上説明したように、本発明によれば、CV・VCやVCVのように音声素片内のパワー変動が大きくなるような音韻単位を波形編集の単位としても適切なパワー制御を行うことができ、高品位な合成音声を生成することができる。

## 【図面の簡単な説明】

## 【図 1】

第 1 の実施形態による音声合成装置のハードウェア構成を示すブロック図である。

【図 2】

本実施形態による音声合成処理の手順を示すフローチャートである。

【図 3】

部分素片のパワー推定に必要な要因の例を示す図である。

【図 4】

部分素片が有声音の素片か無声音の素片かを判定するために参照するテーブルのデータ構成例を示す図である。

【図 5】

有声音パワー推定用に学習された数量化 I 類の係数表の例を示す図である。

【図 6】

無声音パワー推定用に学習された数量化 I 類の係数表の例を示す図である。

【図 7】

第 2 の実施形態による音声合成処理の手順を説明するフローチャートである。

【図 8】

第 3 の実施形態による素片辞書作成の処理手順を説明するフローチャートである。

【図 9】

図 8 のフローチャートによる素片辞書作成処理を説明する図である。

【図 1 0】

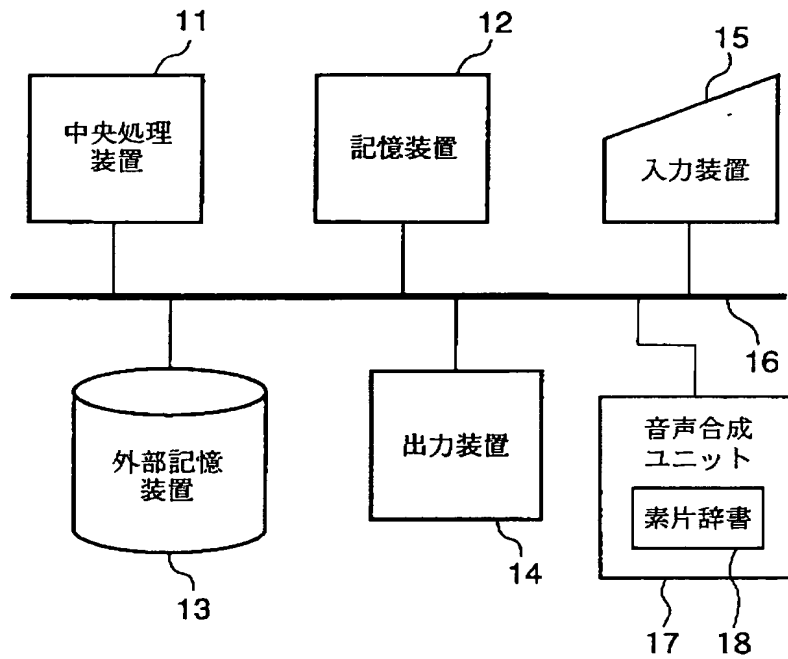
音声素片単位である CV・VC、VCV を説明する図である。

【図 1 1】

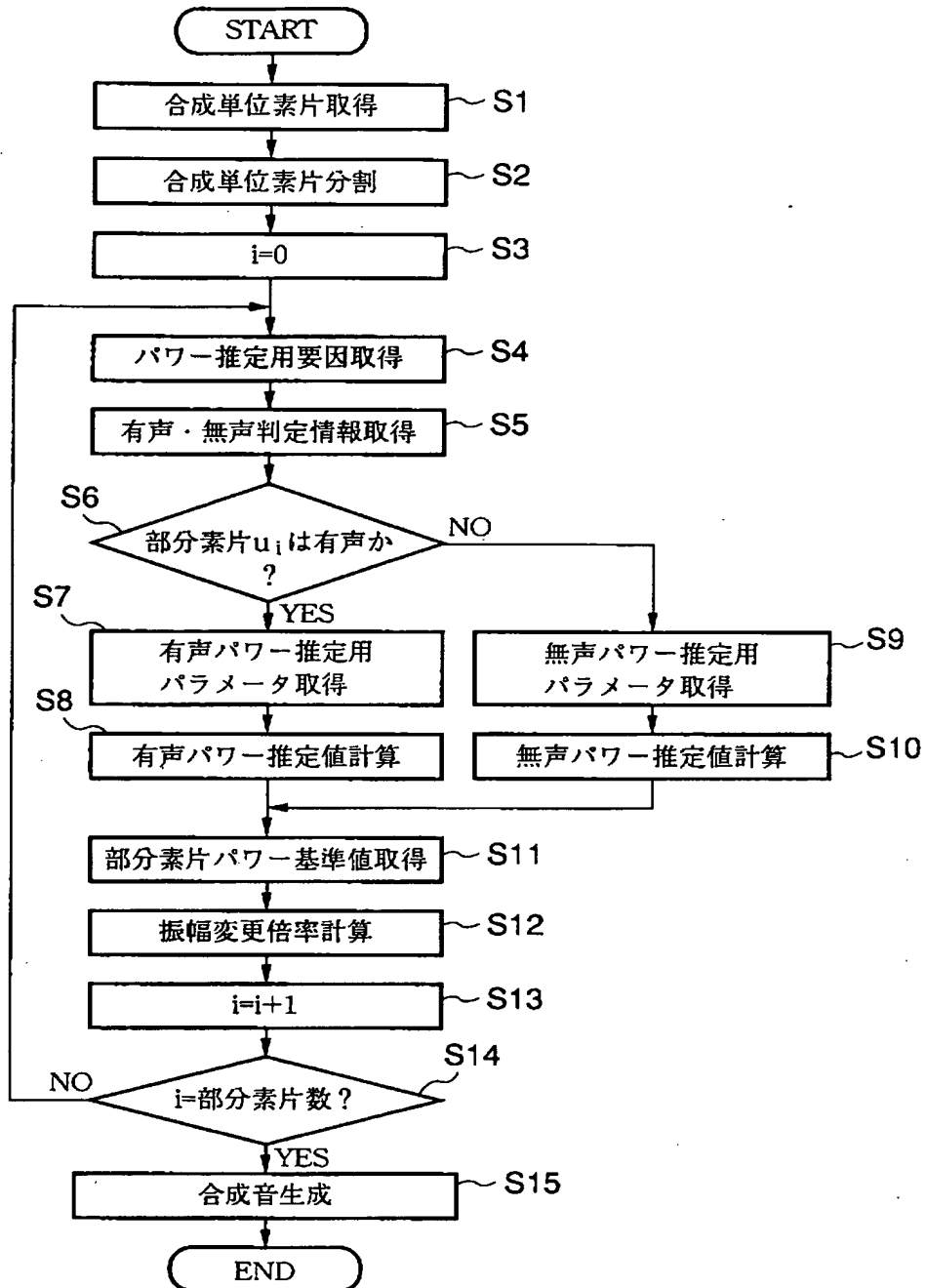
音声波形を微細素片に分割する方法を模式的に示した図である。

【書類名】 図面

【図 1】



【図 2】



【図 3】

要因名	音素種別	モーラ数	アクセント型	モーラ位置	...
値	/i/	5	2	3	...

【図 4】

素片ID	0			1			...
部分素片番号	0	1	2	0	1	2	...
有声無声フラグ	無	有	無	無	有	有	...

【図5】

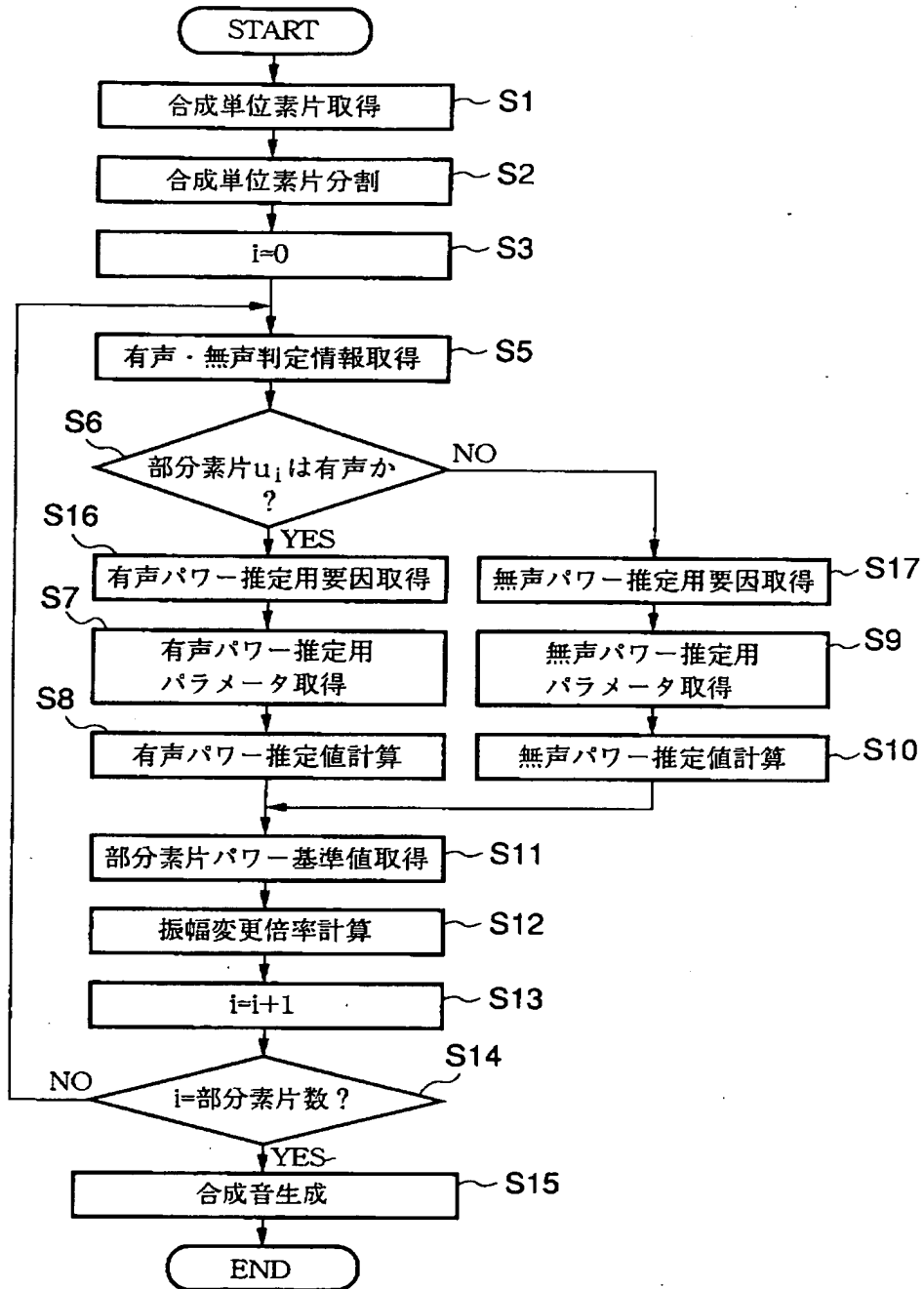
要因名	音素種別			モーラ数			アクセント型			モーラ位置			...
	/a/	/i/	...	1	2	...	0	1	...	1	2	...	
要因													...
係数	21730	10492	...	-1320	-4174	...	236	-10	...	8121	5230	...	...



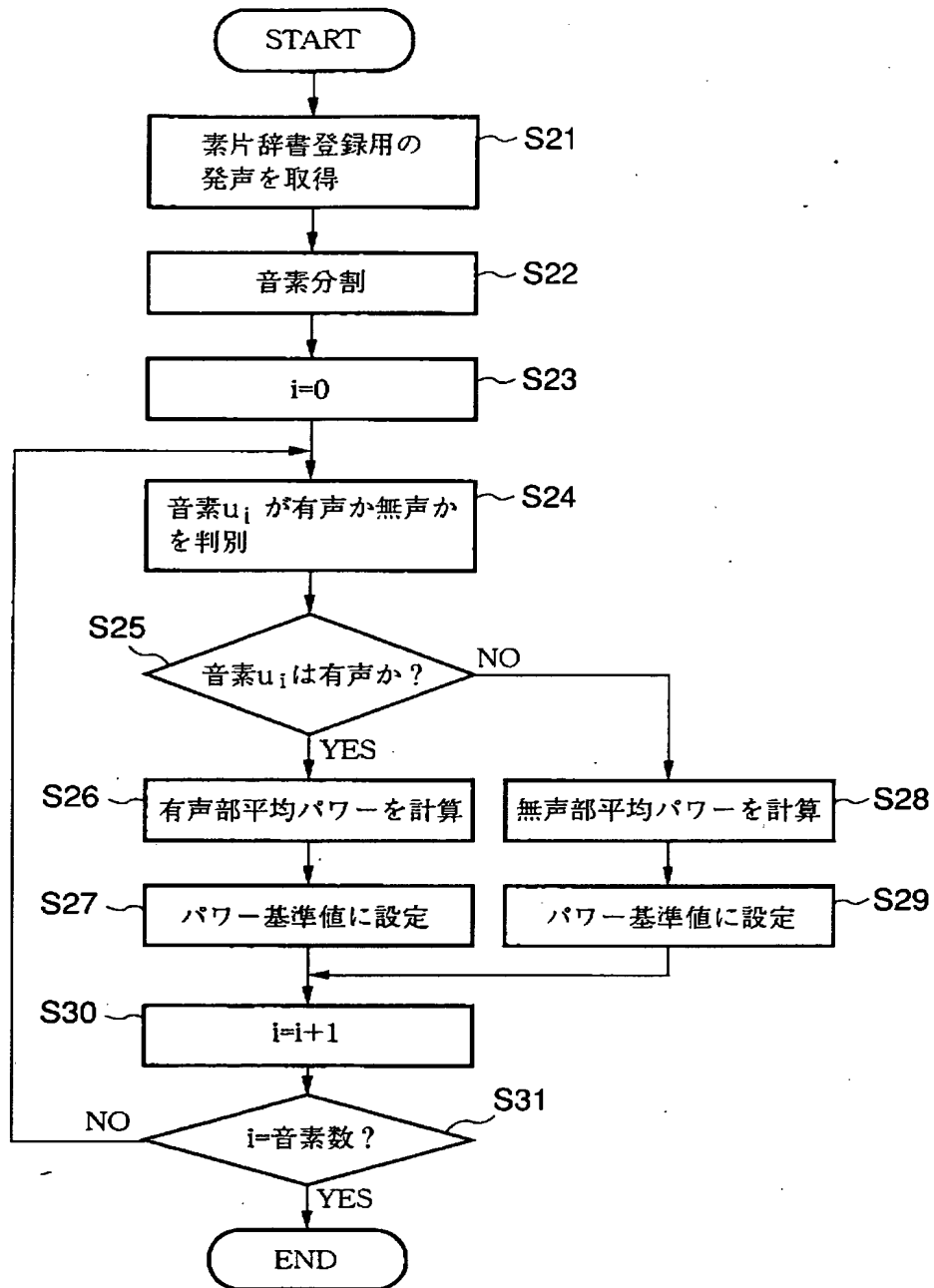
【図 6】

要因名	音素種別		モーラ数			アクセント型			モーラ位置			...
	/a/	/i/	1	2	...	0	1	...	1	2	...	
要因												...
係数	1030	985	-20	-45	...	-4	-17	...	945	753	...	...

【図 7】



【図 8】

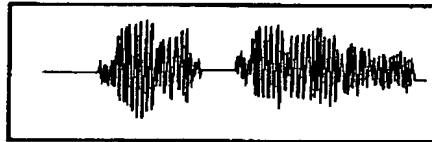


【図 9】

(a) 仮名表記

タ	カ	イ
---	---	---

(b) 音声波形



(c) 音素

/t/	/a/	/k/	/a/	/i/
-----	-----	-----	-----	-----

(d) 有声無声  
判定結果

U	V	U	V	V
---	---	---	---	---

(e) 音素パワー  
(基準値)

893	21473	1204	34820	17201
-----	-------	------	-------	-------

(f) 部分素片



(g) CV・VC

-t	t.a	a.k	k.a	a.i	i.-
----	-----	-----	-----	-----	-----

【図 10】

仮名表記

タ	カ	イ
---	---	---

音素

/t/	/a/	/k/	/a/	/i/
-----	-----	-----	-----	-----

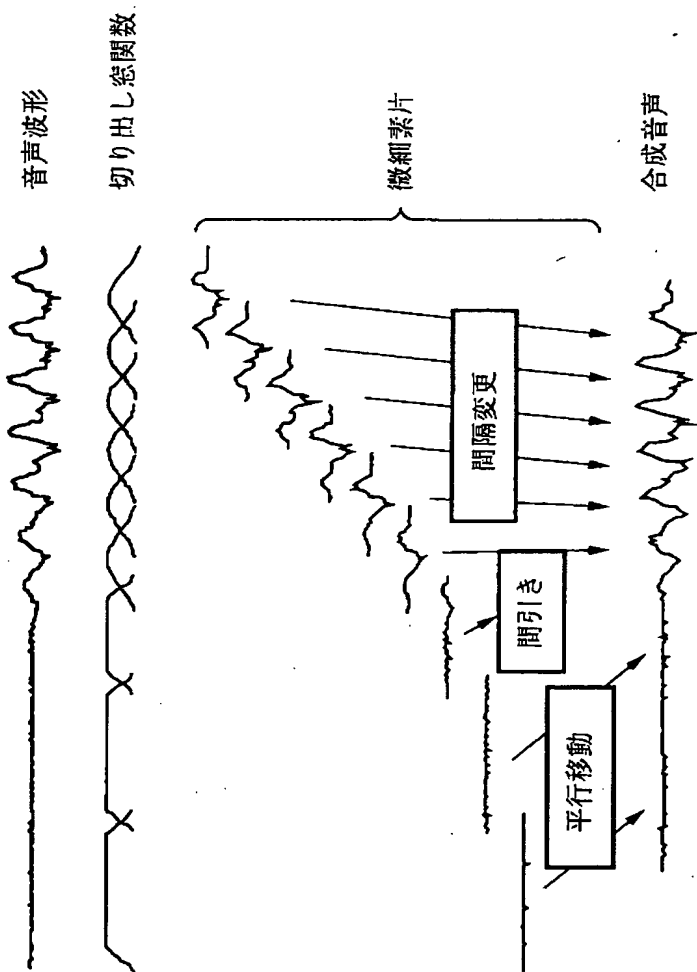
CV・VC

-t	t.a	a.k	k.a	a.i	i.-
----	-----	-----	-----	-----	-----

VCV

-t.a	a.k.a	a.i.-
------	-------	-------

【図 11】



【書類名】 要約書

【要約】

【課題】 音声素片内のパワー変動が大きくなるような音韻単位を波形編集の単位としても適切なパワー制御を行う。

【解決手段】 音声合成単位に分割された合成単位素片を取得し、取得された合成単位素片を音素境界で分割して部分素片を取得する（ステップ S 1、S 2）。そして、取得された部分素片の各々に関してパワー値を推定し（ステップ S 4～S 10）、部分素片の各々のパワー値を推定されたパワー値に変更するための振幅倍率を得る（ステップ S 11～S 12）。全ての部分素片について振幅倍率を得て、これに基づいて合成単位素片の振幅を変更して合成音を生成する（ステップ S 14、S 15）ことによりパワー制御を実行する。

【選択図】 図 2

出 願 人 履 歴 情 報

識別番号 [000001007]

1. 変更年月日 1990年 8月30日  
[変更理由] 新規登録  
住 所 東京都大田区下丸子3丁目30番2号  
氏 名 キヤノン株式会社